

# VU Research Portal

## Levels of evidence

de Vet, H.C.W.; van Tulder, M.W.; Bouter, L.M.

### **published in**

Journal of Clinical Epidemiology  
2003

### **DOI (link to publisher)**

[10.1016/S0895-4356\(03\)00196-3](https://doi.org/10.1016/S0895-4356(03)00196-3)

### **document version**

Publisher's PDF, also known as Version of record

## [Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

de Vet, H. C. W., van Tulder, M. W., & Bouter, L. M. (2003). Levels of evidence: intellectual aid or absolute judgement? *Journal of Clinical Epidemiology*, 56(9), 917-918. [https://doi.org/10.1016/S0895-4356\(03\)00196-3](https://doi.org/10.1016/S0895-4356(03)00196-3)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## Levels of evidence Intellectual aid or absolute judgement?

Henrica C.W. de Vet\*, Maurits W. van Tulder, Lex M. Bouter

*Institute for Research in Extramural Medicine, VU University Medical Center, Van der Boeorchestraat 7 1081 BT Amsterdam, The Netherlands*

Accepted 12 March 2003

In systematic reviews statistical pooling is not always possible due to inadequate reporting of the results of original studies. In these reviews, a qualitative analysis using levels of evidence may be performed to summarize the evidence and to formulate conclusions [1]. This is explicit and reproducible, because it explains the exact meaning of labels like strong, moderate, and limited evidence. In the course of time, different sets of levels of evidence have been published [2–5]. All these sets are arbitrary and based on common sense at best. Ferreira et al. [6] show in their article that these different criteria may lead to different conclusions. They advise readers to be cautious when interpreting conclusions of systematic reviews that use levels of evidence. We fully agree, but not merely because different sets of criteria exist. If only one set of levels of evidence would exist this warning would even be more necessary. The use of levels of evidence is essentially an arbitrary and subjective way of summarizing evidence. Typically levels of evidence take into account the quality of the studies and the consistency of the results. In contrast to meta-analysis (statistical pooling), levels of evidence do not include the size of the effect. In the interpretation of both the conclusions of a study and its methodologic quality, there is some subjectivity involved. For example, grading the conclusions is difficult when there is borderline statistical significance, when a positive effect is observed for only part of many outcome measures studied, or when the reviewers do not agree with the authors' conclusions.

A similar example of arbitrary definitions is the interpretation of kappa values. Ferreira et al. [6] use the benchmarks proposed by Landis and Koch [7] to interpret their kappa values. Other often-used benchmarks are the ones proposed by Fleiss [8], which would have changed the conclusions about the disagreements between different sets of levels of evidence in Ferreira's paper [6] slightly. Existence of different rating systems emphasizes their subjectivity. Kappa

values are dependent on a number of factors [9], and an absolute interpretation would give false certainty. Another illustrative example in the field of systematic reviews is the existence of many different instruments to measure the methodologic quality of a randomized clinical trial. Jüni et al. [10] applied 25 different quality instruments to the same review and found that the conclusion of the review was dependent on the quality instrument that was used. Of course, in a review with convincing and indisputable results the conclusion will be the same regardless of the quality instruments and the set of levels of evidence used. Sensitivity analysis to study the robustness of the conclusions of a systematic review, as both Ferreira et al. [6] and Jüni et al. [10] suggest, is an advisable solution.

The most striking observation in the article of Ferreira et al. [6] is that the different levels of evidence came from the same research institute [1,2–4]. This illustrates how arbitrary the levels of evidence are, and the difficulty we had in reaching consensus. Recently, we proposed levels of evidence for reviews with incomplete statistical pooling, but also these are arbitrary [11].

An interesting question is whether levels of evidence should be the same for all situations. Less stringent rules for evidence could be proposed for situations in which there is an urgent need for an effective therapy and very stringent rules in fields where an abundance of effective therapies exists. In these latter fields one wants to be quite sure whether an additional therapy is really more effective than all existing ones.

In our opinion, the existence of different sets of levels of evidence is not alarming, and emphasizes the essential relativity. However, we recognize the disadvantage that reviewers may be tempted to choose the set of criteria that gives them the most favorable conclusions. Therefore, within the field of low back pain the Cochrane Back Review Group has recently acknowledged the need for standardization and included a proposal for a specific rating system in the update of the method guidelines for systematic reviews in the field of back pain [12].

\* Corresponding author. Tel.: +31-20-444-8176; fax: +31-20-444-8181.

E-mail address: [HCW.de\\_Vet.emgo@med.vu.nl](mailto:HCW.de_Vet.emgo@med.vu.nl) (H.C.W. de Vet).

## References

- [1] Van Tulder M, Koes BW, Bouter LM. Conservative treatment of acute and chronic non specific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997;22:2128–56.
- [2] Van Tulder MW, Ostelo R, Vlaeyen JWS, Linton SJ, Morley SJ, Assendelft WJJ. Behavioural treatment for chronic low back pain: a systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine* 2000;25:2688–99.
- [3] Van Poppel MNM, Koes BW, Smid T, Bouter LM. A systematic review of controlled clinical trials on the prevention of low back pain in industry. *Occup Environ Med* 1997;54:841–47.
- [4] Van der Windt DAWN, Van der Heijden GJMG, Van der Berg SGM, Ter Riet G, De Winter AF, Bouter LM. Ultrasound therapy for musculoskeletal disorders: a systematic review. *Pain* 1999;81:257–71.
- [5] Berghmans LCM, Hendriks HJM, Bo K, Hay -Smith EJ, De Bie RA, Van Doorn ESCVW. Conservative treatment of stress urinary incontinence in women: a systematic review of randomised clinical trials. *Br J Urol* 1998;82:181–91.
- [6] Ferreira PH, Ferreira ML, Maher CG, Refshauge K, Herbert RD, Latimer J. Effect of applying different “levels of evidence” criteria on conclusions of Cochrane reviews of interventions for low back pain. *J Clin Epidemiol* 2002;55:1126–9.
- [7] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1997;33:159–74.
- [8] Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986.
- [9] Feinstein AR, Cicchetti D. High agreement but low kappa: I. The problem of two paradoxes. *J Clin Epidemiol* 1990;43:543–9.
- [10] Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
- [11] Smidt N, Assendelft WJJ, van der Windt DAWM, Hay EM, Buchbinder R, Bouter LM. Corticosteroid injections for lateral epicondylitis: a systematic review. *Pain* 2002;96:23–40.
- [12] Van Tulder MW, Furlan A, Bombardier C, Bouter LM, and the Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290–9.